

How can I perform quantile normalization in R?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform quantile normalization in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165011>

Quantile normalization is a statistical method used to standardize the distribution of data, specifically in cases where the data may be initially skewed or have different scales. This technique is commonly used in data preprocessing for various statistical analyses, such as gene expression studies. In R, quantile normalization can be performed using the "normalize.quantiles" function from the "preprocessCore" package. This function takes in a matrix or data frame of numerical values and adjusts the distribution to match a desired theoretical distribution, typically a uniform distribution. The resulting normalized data can then be used for further analysis.

Perform Quantile Normalization in R

In statistics, is a method that makes two distributions identical in statistical properties.

The following example shows how to perform quantile normalization in R.

Example: Quantile Normalization in R

Suppose we create the following data frame in R that contains two columns:

```
#make this example reproducible  
set.seed(0)
```

```
#create data frame with two columns  
df <- data.frame(x=rnorm(1000),  
y=rnorm(1000))
```

```
#view first six rows of data frame
```

head(df)

x y

```
1 1.2629543 -0.28685156
2 -0.3262334 1.84110689
3 1.3297993 -0.15676431
4 1.2724293 -1.38980264
5 0.4146414 -1.47310399
6 -1.5399500 -0.06951893
```

We can use the `quantile` and `apply` functions to calculate the quantiles for both x and y:

```
#calculate quantiles for x and y
apply(df, function(x) quantile(x, probs = seq(0, 1, 1/4)))
```

x y

```
0% -3.23638573 -3.04536393
25% -0.70845589 -0.73331907
50% -0.05887078 -0.03181533
75% 0.68763873 0.71755969
100% 3.26641452 3.03903341
```

Notice that x and y have similar values for the quantiles, but not identical values.

For example, the value at the 25th percentile for x is -0.708 and the value at the 25th percentile for y is -0.7333.

To perform quantile normalization, we can use the `normalize.quantiles()` function from the package in R:

```
library(preprocessCore)
```

```
#perform quantile normalization
```

```
df_norm <-
```

```
as.data.frame(normalize.quantiles(as.matrix(df)))
```

```
#rename data frame columns
```

```
names(df_norm) <- c('x', 'y')
```

```
#view first six row of new data frame
```

```
head(df_norm)
```

```
x y
```

```
1 1.2632137 -0.28520228
```

```
2 -0.3469744 1.82440519
```

```
3 1.3465807 -0.16471644
```

```
4 1.2692599 -1.34472394
```

```
5 0.4161133 -1.43717759
```

```
6 -1.6269731 -0.07906793
```

We can then use the following code to calculate the quantiles for both x and y again:

```
#calculate quantiles for x and y  
sapply(df_norm, function(x) quantile(x, probs = seq(0, 1,  
1/4)))
```

x y

```
0% -3.14087483 -3.14087483  
25% -0.72088748 -0.72088748  
50% -0.04534305 -0.04534305  
75% 0.70259921 0.70259921  
100% 3.15272396 3.15272396
```

Notice that the quantiles are identical for x and y now.

We would say that x and y have been quantile normalized. That is, the two distributions are now identical in statistical properties.

Additional Resources

The following tutorials explain how to perform other common tasks in R: